

ZOMATO DATA ANALYSIS USING PYTHON





OVERVIEW:

This project involves analyzing a dataset from Zomato, one of the leading food delivery and restaurant discovery platforms. The dataset contains key information about various restaurants including their names, availability of online orders and table booking, user votes, ratings, cost for two people, and listing types. Using Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn, the data was explored, cleaned, and visualized to extract actionable insights and understand customer preferences and restaurant trends.





OBJECTIVE:

- ▶ Analyze restaurant behavior, customer engagement, and service trends using Zomato data.
- ▶ Identify patterns in online ordering, table booking, and pricing.
- ▶ Use Python libraries to clean, explore, and visualize the data.





Q1. How many restaurants allow online orders?

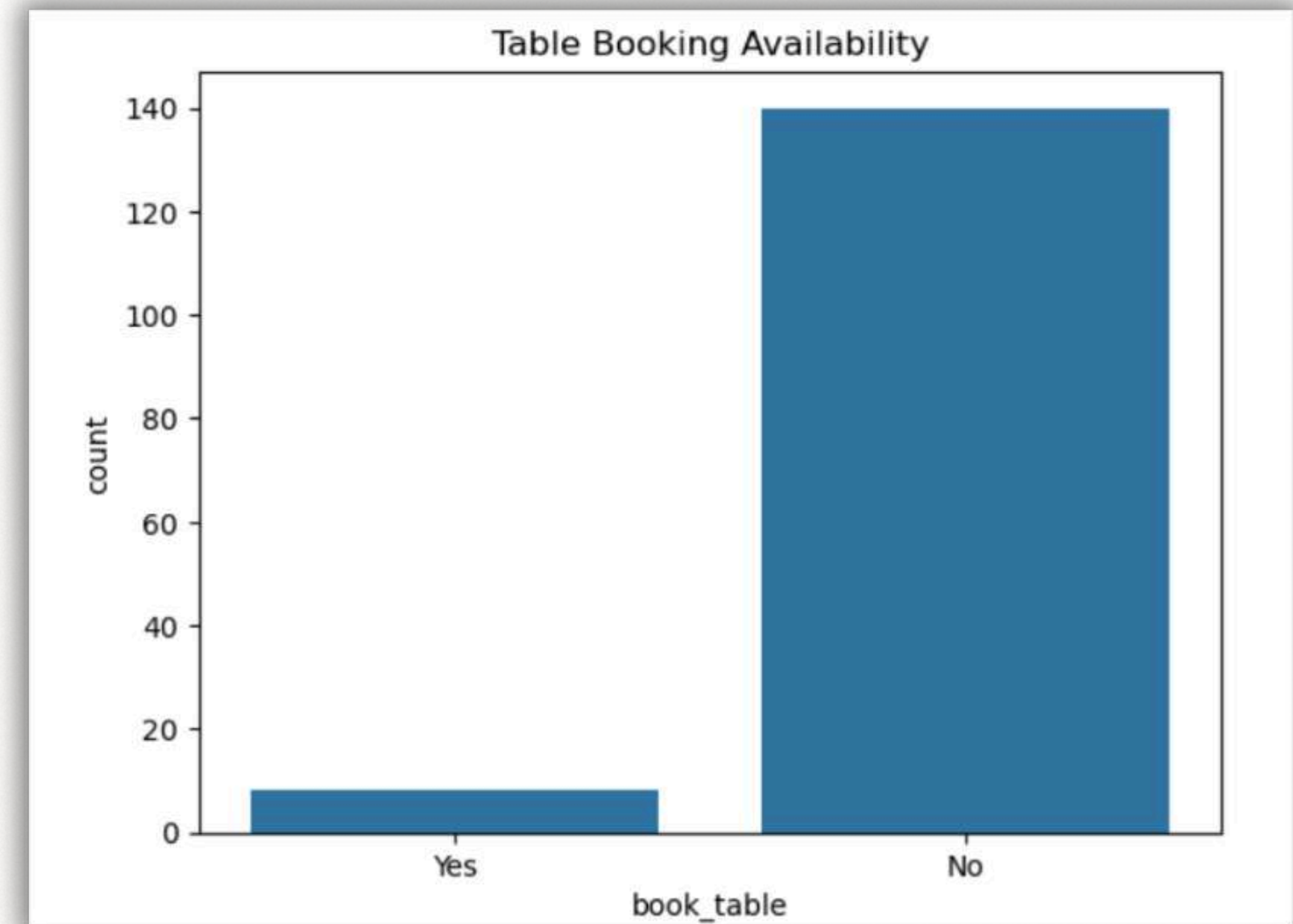
```
sns.countplot(x='online_order', data=dataframe)
plt.title("Online Ordering Availability")
plt.show()
```





Q2. How many restaurants offer table booking?

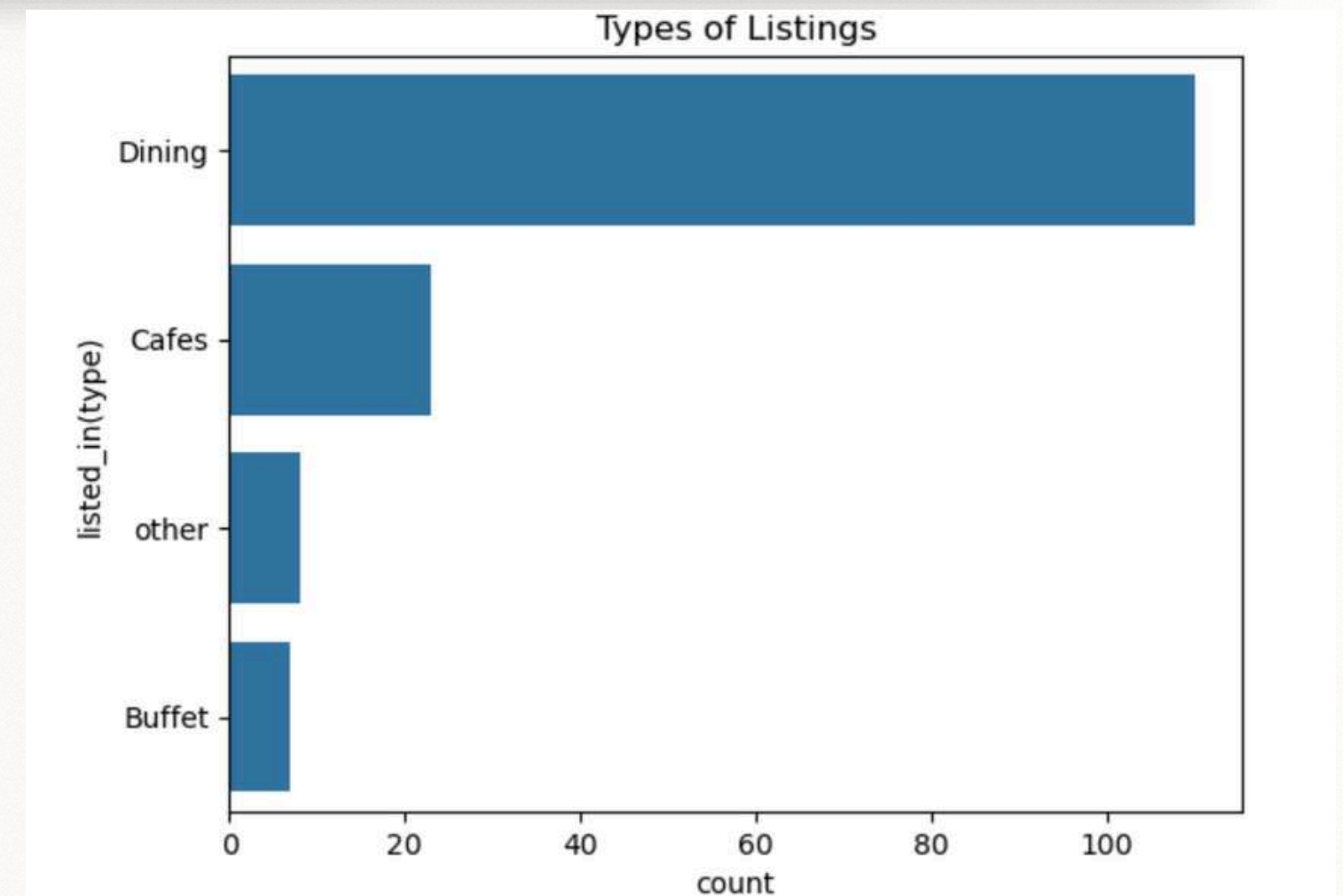
```
sns.countplot(x='book_table', data=dataframe)
plt.title("Table Booking Availability")
plt.show()
```





```
sns.countplot(y='listed_in(type)', data=dataframe, order=dataframe['listed_in(type)'].value_counts().index)
plt.title("Types of Listings")
plt.show()
```

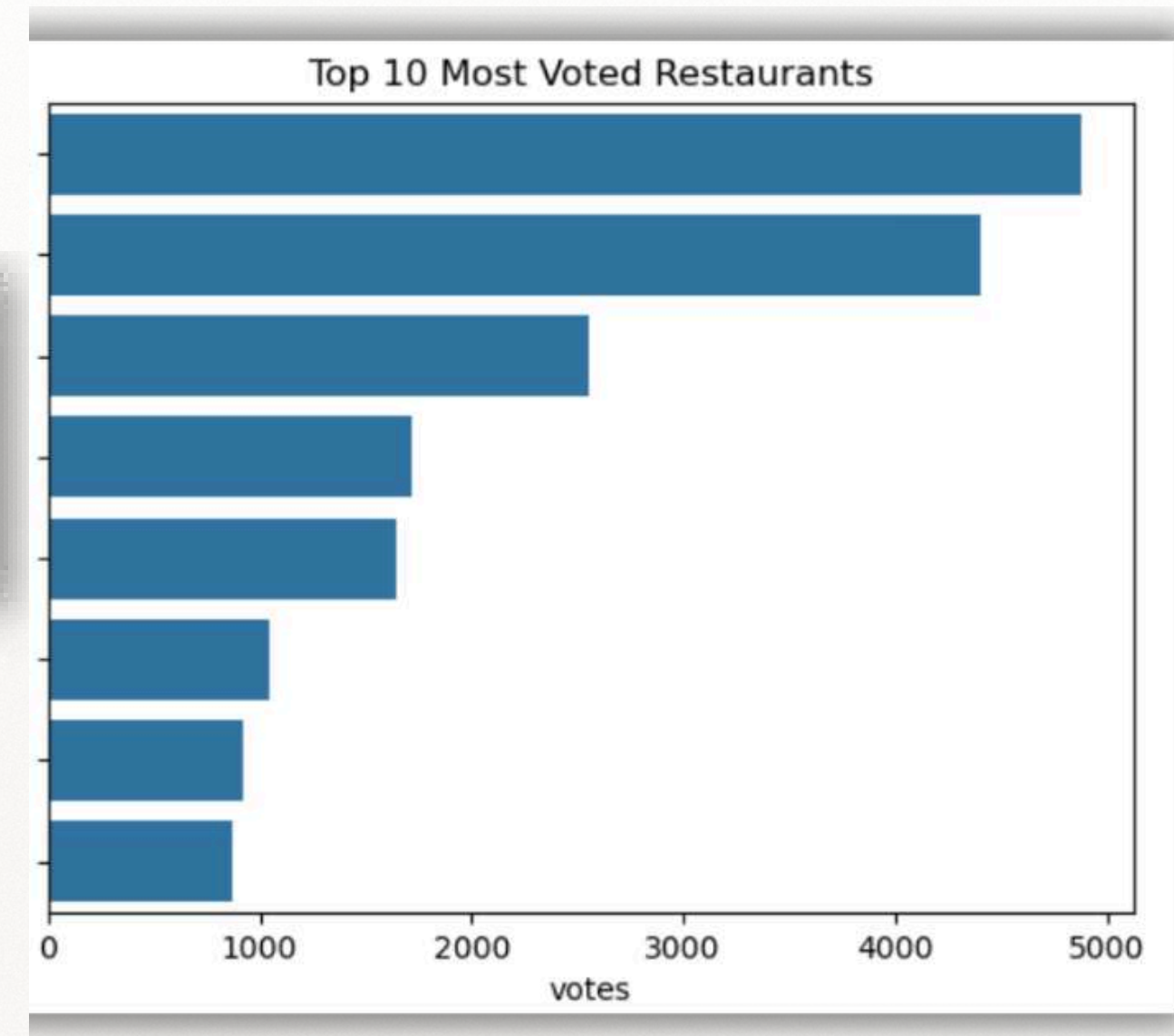
Q3. Which type of listing (e.g., Buffet, Delivery) is most common?





Q 4. What are the top 10 most voted restaurants?

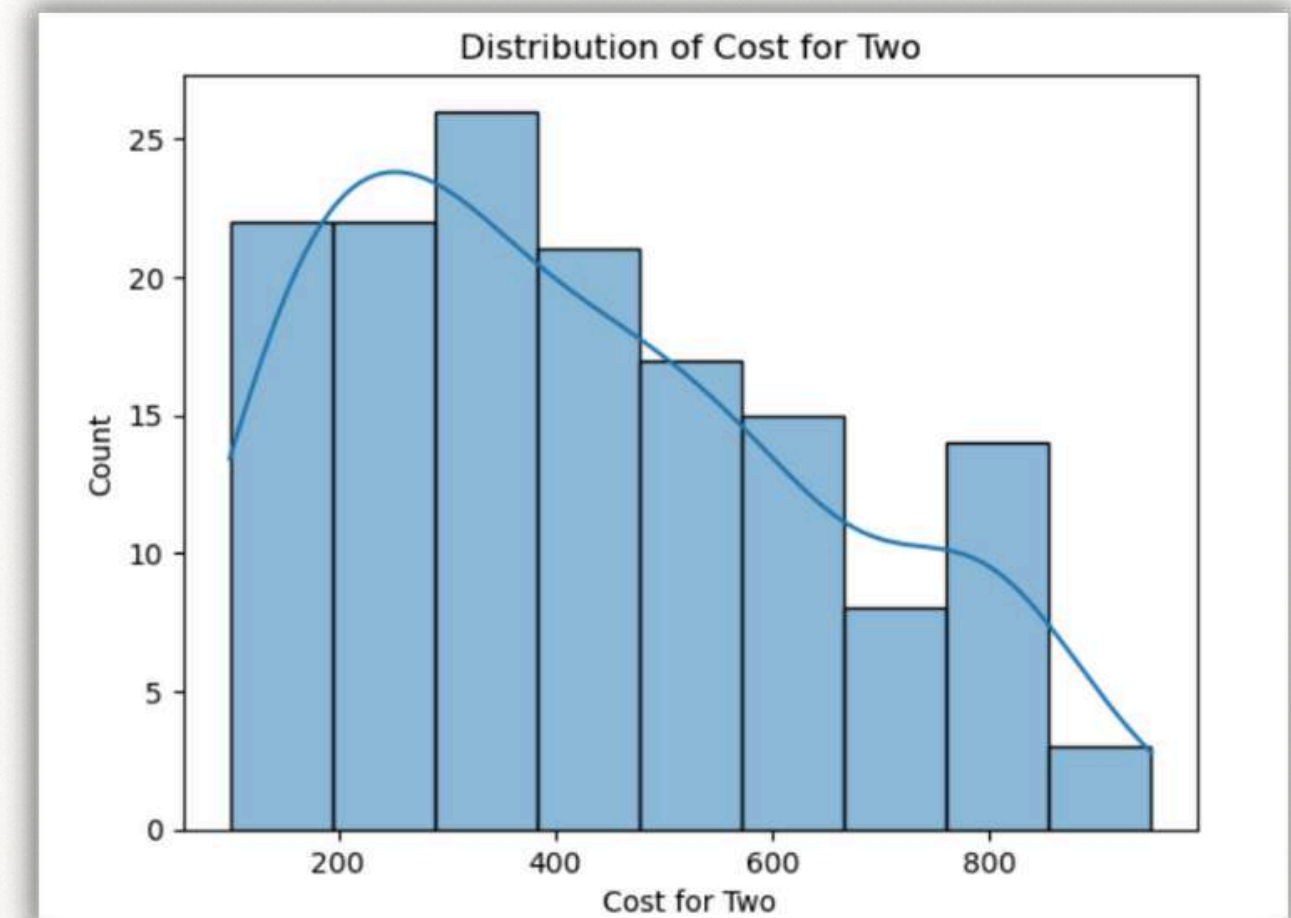
```
top_votes = dataframe.sort_values(by='votes', ascending=False).head(10)
sns.barplot(x='votes', y='name', data=top_votes)
plt.title("Top 10 Most Voted Restaurants")
plt.show()
```





Q5. Distribution of Approximate Cost for Two People.

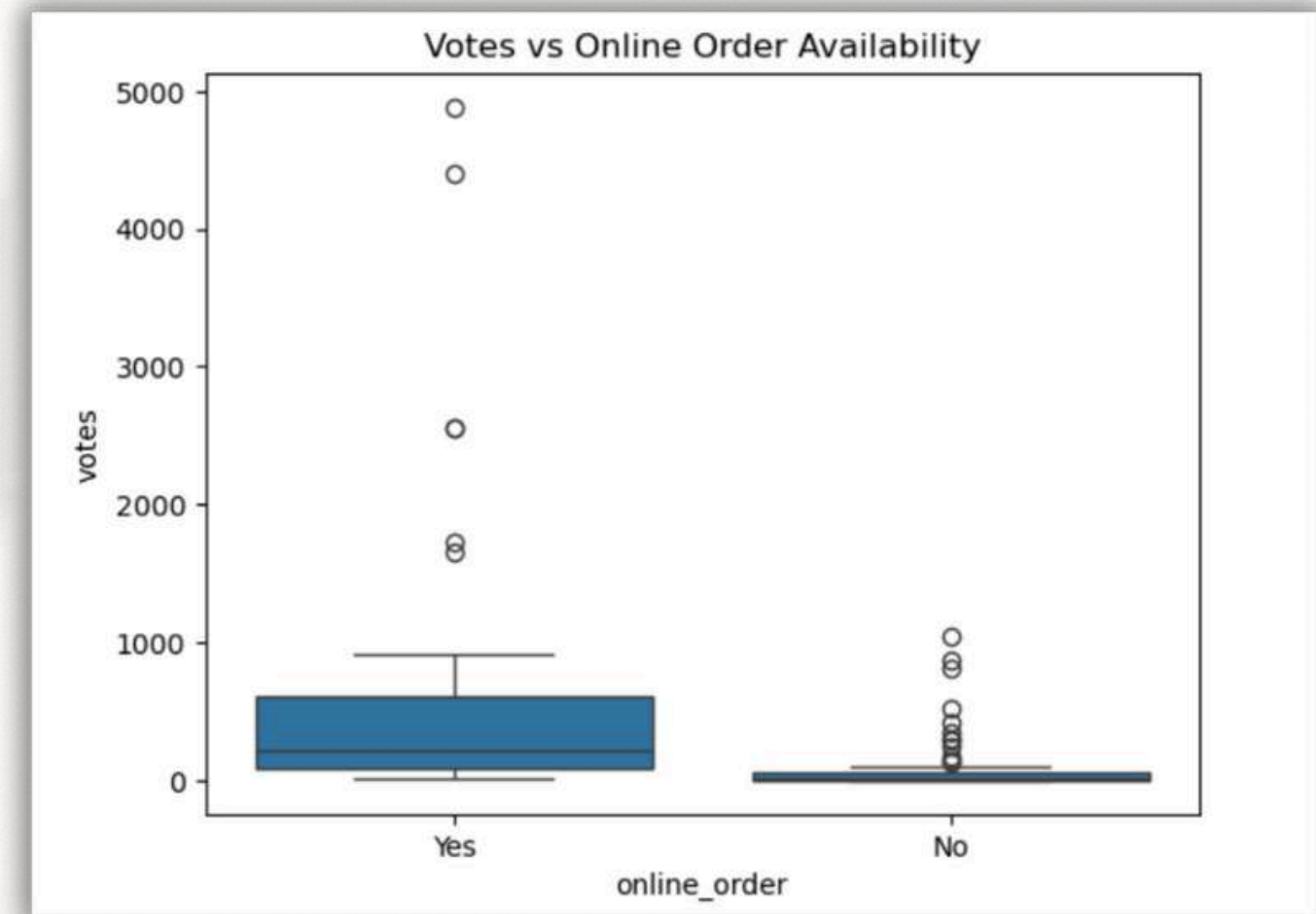
```
sns.histplot(dataframe['approx_cost(for two people)'], kde=True)
plt.title("Distribution of Cost for Two")
plt.xlabel("Cost for Two")
plt.show()
```





Q 6. What is the relationship between online order availability and votes?

```
sns.boxplot(x='online_order', y='votes', data=dataframe)
plt.title("Votes vs Online Order Availability")
plt.show()
```



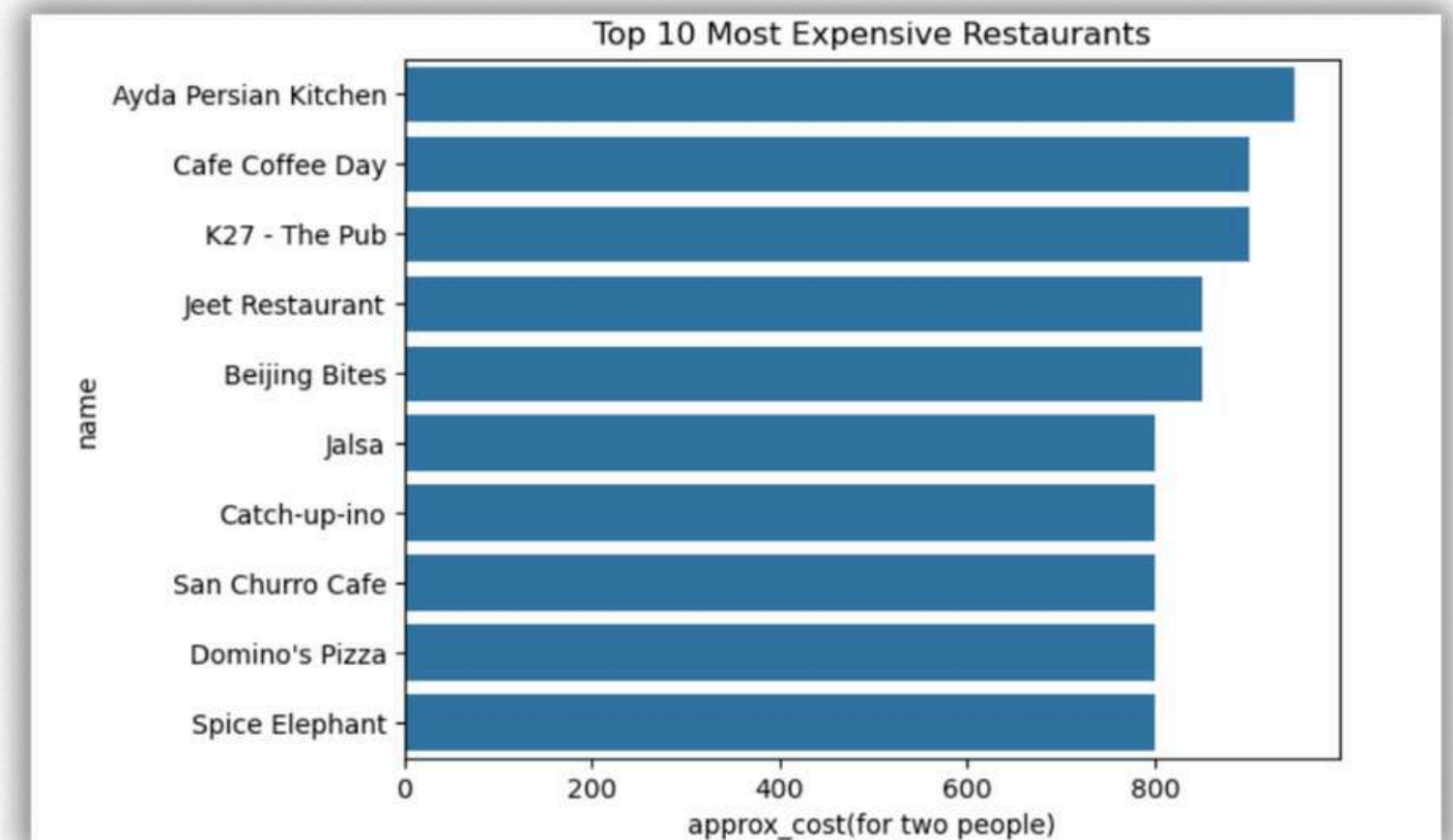
AG



QUESTIONS:

```
top_cost = dataframe.sort_values(by='approx_cost(for two people)', ascending=False).head(10)
sns.barplot(x='approx_cost(for two people)', y='name', data=top_cost)
plt.title("Top 10 Most Expensive Restaurants")
plt.show()
```

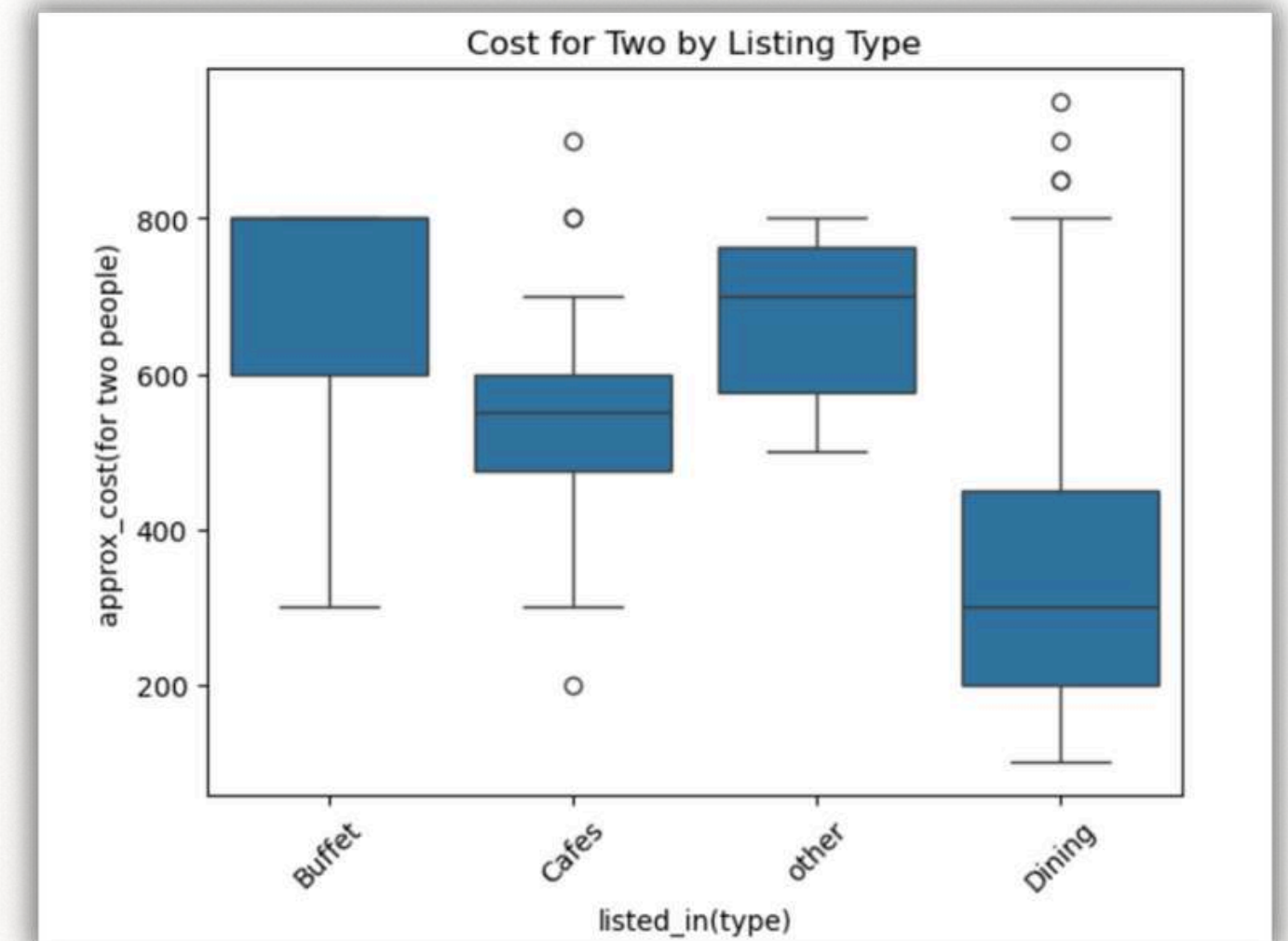
Q7. Which restaurants have the highest average cost for two people?





Q8. Comparison of cost distribution based on listing type.

```
sns.boxplot(x='listed_in(type)', y='approx_cost(for two people)', data=dataframe)
plt.title("Cost for Two by Listing Type")
plt.xticks(rotation=45)
plt.show()
```



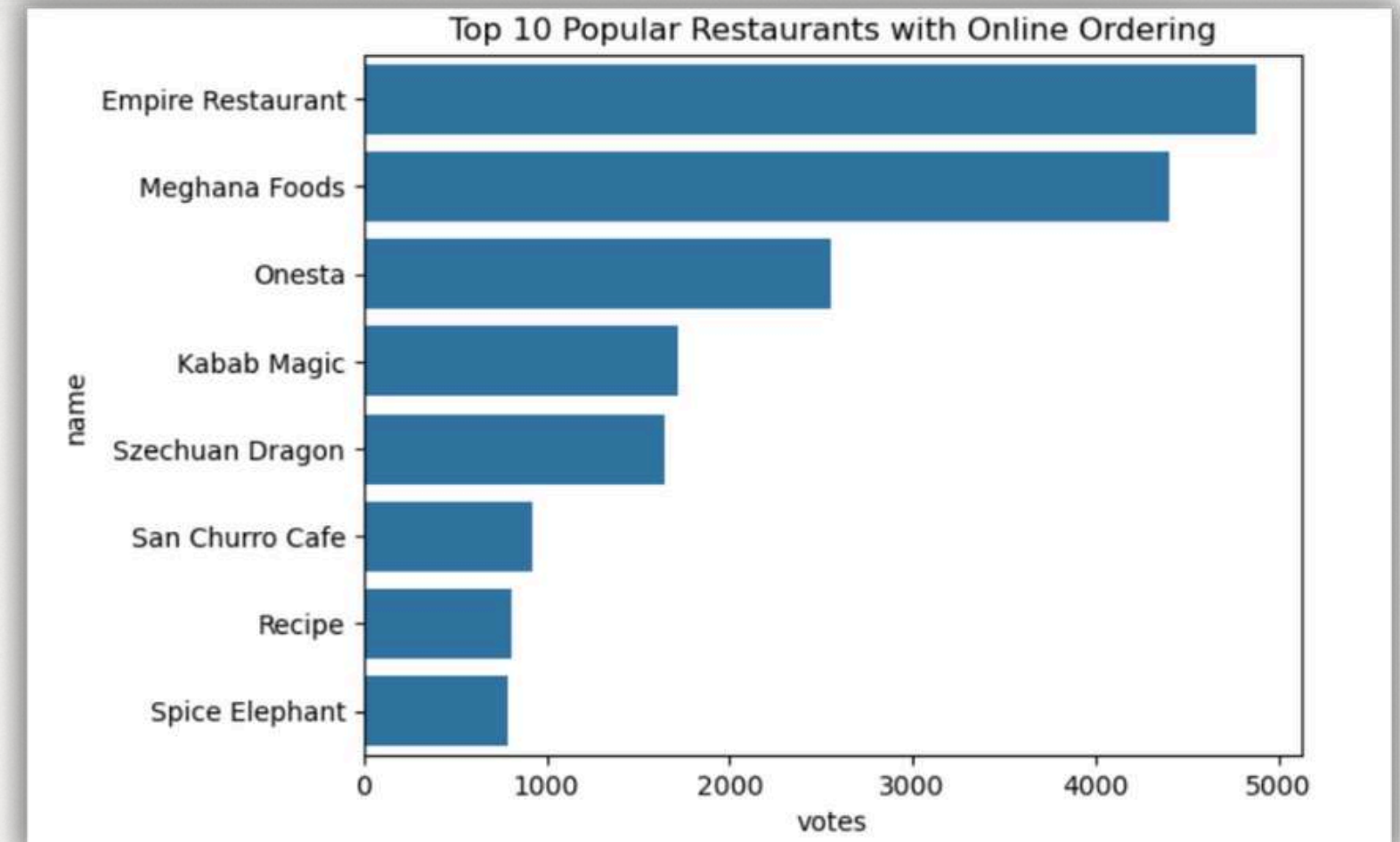
AG



QUESTIONS:

```
popular = dataframe[dataframe['online_order'] == 'Yes'].sort_values(by='votes', ascending=False).head(10)
sns.barplot(x='votes', y='name', data=popular)
plt.title("Top 10 Popular Restaurants with Online Ordering")
plt.show()
```

Q9. Top 10 most popular restaurants (based on votes and online order availability) .

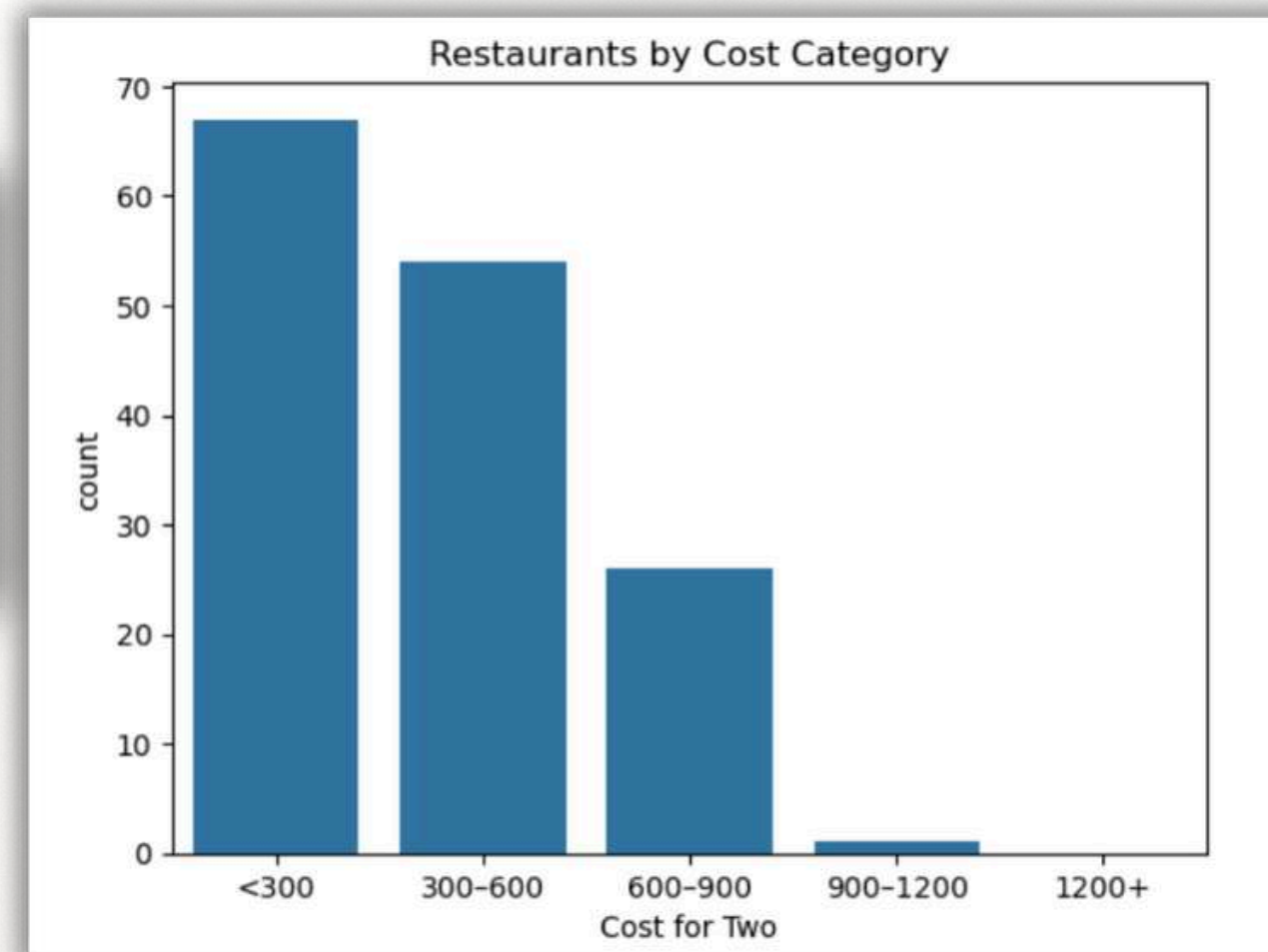




QUESTIONS:

Q10. How many restaurants fall into each cost category (binned)?

```
bins = [0, 300, 600, 900, 1200, 1500]
labels = ['<300', '300-600', '600-900', '900-1200', '1200+']
dataframe['cost_category'] = pd.cut(dataframe['approx_cost(for two people)'], bins=bins, labels=labels)
sns.countplot(x='cost_category', data=dataframe)
plt.title("Restaurants by Cost Category")
plt.xlabel("Cost for Two")
plt.show()
```





CONCLUSION:

- ▶ **Online ordering is more popular than table booking, with the "Delivery" listing type being the most common and widely adopted by restaurants.**
- ▶ **Mid-range pricing and online availability drive higher user engagement, emphasizing the importance of digital services and data-driven decisions in the restaurant industry.**

zomato

Statistics, Facts, & User Counts



THANK YOU